Acta Geodyn. Geomater., Vol. 9, No. 4 (168), 511–520, 2012

DETERMINATION OF CHLORITE, MUSCOVITE, ALBITE AND QUARTZ IN CLAYSTONES AND CLAY SHALES BY INFRARED SPECTROSCOPY AND PARTIAL LEAST-SQUARES REGRESSION

Michal RITZ [1] *, Lenka VACULÍKOVÁ [2], Eva PLEVOVÁ [2], Dalibor MATÝSEK[1] and Jiří MALIŠ [1]

[1] *VŠB-Technical University Ostrava, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic*
[2] *Institute of Geonics of the AS CR, Studentská 1768, 708 00 Ostrava-Poruba, Czech Republic*
**Corresponding author's e-mail:* michal.ritz@vsb.cz

(*Received* April 2012*, accepted* September 2012)

ABSTRACT
The objective of this work is the chemometric quantification of minerals in rocks. A chemometric method was developed for the determination of chlorite, muscovite, albite and quartz in claystones and clay shales using infrared spectroscopy. Bromide pellets and diffuse reflectance were used to measure the infrared spectra; principal component analysis and partial least-squares regression were used as chemometric methods. Spectral regions (4000-3000 cm^{-1} and 1300-400 cm^{-1}) containing important spectral information were chosen by principal component analysis. The calibration models were created by a partial least-squares regression. The mean relative error and relative standard deviation were calculated for the assessment of accuracy and reproducibility. The value of the mean relative error was about 10 % for most of the calibration models. The value of the relative standard deviation ranged from 1.1 to 3.0 % for most calibration models based on diffuse reflectance spectra and from 4.0 to 9.2 % for most calibration models based on spectra obtained with bromide pellets.

KEYWORDS: claystone and clay shale, infrared spectroscopy, chemometrics, chlorite, muscovite, albite, quartz

INTRODUCTION

The type and content of minerals present in rocks have a significant influence on the behavior and properties of the rocks as well as on the whole rock massif. A detailed qualitative and quantitative mineral analysis is therefore a necessary step for the characterization of the properties of the rocks in geological, geochemical and geomechanical studies. The experimental results can be applied subsequently in geomechanics and mining activities.

Several existing conventional analytical methods can be used to examine the mineral composition of rocks: optical microscopy, electron microscopy, X-ray diffraction (XRD), infrared (IR) spectroscopy, Raman spectroscopy, thermal gravimetric/differential thermal analysis (TG/DTA) and bulk chemistry analysis (Kodama et al., 1989; Chipera and Bish, 2001; Srodon, 2002; Vogt et al., 2002). Unfortunately, the exact determination of minerals (especially clay minerals) in the rocks by these methods is rather complicated and often inaccurate. The main analytical difficulties are related to the variable chemical composition and common structural anomalies of clay minerals. The individual clay minerals occur in the form of mixtures with various ratios of the particular clay minerals.

Current IR spectroscopy represents a fast, reliable and efficient tool for phase analysis. This method especially combined with chemometric methods can also be used for the purpose of quantitative analysis. The theory of chemometric methods has been described in many papers and books, such as Fredericks et al., 1985; Geladi and Kowalski, 1986; Lorber and Kowalski, 1988; Martens and Naes, 1989. Numerical methods, such as multiple linear regression (MLR), principal component regression (PCR) or partial least-squares (PLS) regression, are increasingly being used to circumvent the problems posed by the presence of interferences, spectral overlap or major matrix effects (Luis et al., 2004). There are many applications of the chemometric method to IR spectroscopic analysis (Fuller et at., 1988; Haaland and Thomas, 1988; Iñón et al., 2003; Armenta et al., 2007; Breen et al., 2008). In recent years, chemometric methods have also been applied to other analytical methods, such as voltammetry or chromatography (Moneeb, 2006; Al-Degs et al., 2008; Wagieh et al., 2010; Zapata-Urzua et al., 2010).

The spectroscopic applications of chemometric methods tend to use the full spectrum. This approach provides a more accurate description of the model than a single measurement at a specific wavelength or wavenumber, respectively. However, the full-spectrum applications pose some problems: 1) part of the information that is gathered can be redundant; and 2) the measured signal for some wavelengths may be noisy or nonlinear (Luis et al., 2004). The most useful

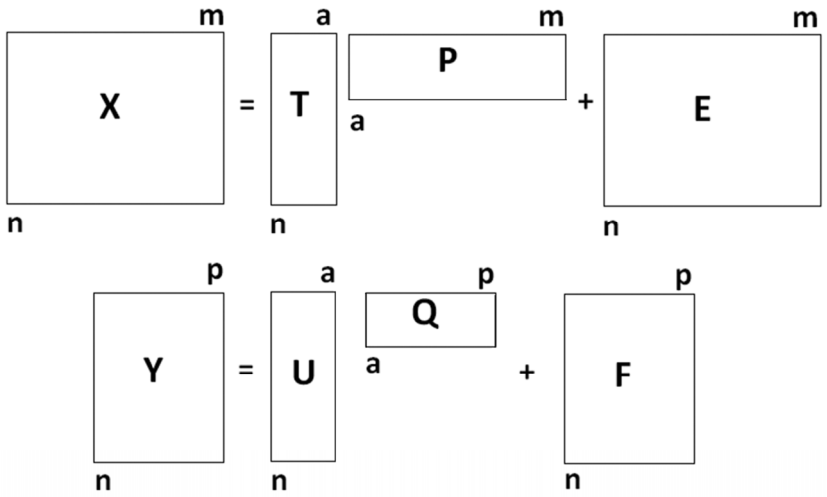

Fig. 1 Graphical representation of PLS (Geladi and Kowalski, 1986).

techniques for chemometric approach in infrared spectroscopy are principal component regression (PCR) and partial least-squares (PLS) regression. An important feature of PCR is fact that only spectral information (e.g. absorbance, Kubelka-Munk unit, etc…) is used for the generation of basis factors. This was not a problem when the concentration information is absolutely accurate. However, in practice the concentration matrix contains error or noise. Another problem of PCR is that collinearity of absorbance data will make the calibration unstable (Hasegawa, 2002). Some potential problems can be solved by using a more stable chemometric method which takes both (absorbance and concentration) matrices into account simultaneously. For this purpose PLS regression can be used as suitable method. In PLS regression absorbance and concentration matrices are used complementarily in a stable calibration.

PLS regression works with two matrices, ***X*** and ***Y***. The ***X*** matrix contains independent variables – the spectral data (e.g. in absorbance unit). The ***Y*** matrix consists of the dependent variables – quantitative (concentration) data. The NIPALS (Nonlinear Iterative Partial Least Squares) algorithm is most often used for creation of PLS models. The PLS models can be considered as consisting of outer relations (***X*** matrix and ***Y*** matrix individually) and an inner relation (linking both matrices) (Geladi and Kowalski, 1986). The outer relations for the ***X*** *a*nd ***Y*** matrices can be written as follows:

$$X = \sum_h t_h p_h + E_x = TP + E \quad (3)$$

$$Y = \sum_h u_h q_h + E_Y = UQ + F \quad (4)$$

where $\boldsymbol{t}_h$ and $\boldsymbol{u}_h$ are score vectors, $\boldsymbol{p}_h$ and $\boldsymbol{q}_h$ are loading vectors; ***T*** and ***U*** are score matrices, ***P*** and ***Q*** are loading matrices, ***E*** and ***F*** are matrices of residuals. Graphically, equations (3) and (4) can be shown as in Figure 1 (Geladi and Kowalski, 1986). In Figure 1, *n* is number of samples, *m* is number of independent variables (e.g. absorbance values), ~~a~~ is number of factors and *p* is number of dependent variables (concentration values). PLS comprises relatively complicated calculation procedures. Their detailed description is beyond the intention of this paper. A more detailed description of a mathematical algorithm of PLS modeling can be found in many books or paper (e.g.: Geladi and Kowalski, 1986; Wold et al., 2001; Hasegawa, 2002).

The PLS regression can be categorized into two procedures. PLS1 (sometimes called standard PLS) and PLS2 (sometimes called global PLS). PLS1 employs information from only one chemical constituent to make the calibration; PLS1method works with the one-column ***Y*** matrix. PLS2 uses two and more chemical constituent simultaneously.

The aim of this study is the determination of majority minerals (chlorite, muscovite, albite and quartz) in claystones and clay shales by IR spectroscopy combined with partial least-squares regression.

EXPERIMENTAL

SAMPLES

Eighty-six samples (B1-B78 and CS1-CS8) of claystones and clay shales were used for this research. The seventy-eight samples (B1-B78) were used as a calibration set and eight samples (CS1-CS8) were used as control samples. In calibration set were thirty-two samples of claystones (B1-B32) and forty-six samples of clay shales (B33-B78). Four control samples (CS1-CS4) were claystones; four control samples (CS5-CS8) were clay shales. The selection of control samples was performed with regard to cover the entire concentration range of the analyzed minerals. All the samples were obtained from the collection of VŠB-Technical University, Ostrava.

Claystone is a compact very fine-grained sedimentary rock consisting primarily of clay-size mineral particles that are commonly represented by clay minerals. Claystones are only partially decomposed in water and their porosity generally varies from 25 to 5 %. Clay shales are similar in composition but do not decompose in the water, and their total porosity is less than 5 %. Moreover, shale is laminated (the rock is made up of many thin layers). Shales that are subject to heat and pressure of metamorphism alter into a hard, fissile (the rock readily splits into thin pieces along the laminations), metamorphic rock known as slate. The samples of claystones were taken from several lower-to-middle Cretaceous strata belonging to the Silesian unit of the Moravian-Silesian Beskydy Mountains. The samples of clay shales are taken from the Kyjovice layers that are stratigraphically adherent to the Lower Carboniferous period of the Moravian-Silesian area.

Every sample was pulverized in an agate mill. Pulverized samples were homogenized by careful shuffling and by repeatedly being spilled.

A list of calibration set including the content of the minerals is shown in Table 1.

POWDER XRD ANALYSIS

The content of minerals in samples of claystones and clay shales was determined by the quantitative reference method - powder XRD analysis. The Rietveld technique, used as the quantification technique for diffraction data (Rietveld, 1969), is the quantitative technique for the crystal structure analysis from powder diffraction data. The theoretical diffractogram is calculated on the basis of structural data (e.g., crystal symmetry, unit cell parameters, atomic coordinates and occupancy) of the minerals that are present. The theoretical diffractogram is subsequently compared with the measured diffraction pattern using multidimensional regression. The Rietveld technique is considered one of the best techniques for the quantification of powder diffraction data, and it has been used in many studies (Chipera and Bish, 2001; Bringley, 1980; Bish and Howard, 1988; Hillier, 2000).

The pulverized and homogenized samples were measured in the cuvettes. Powder diffraction measurements were carried out on a fully-automated diffractometer ID3003 (Rich Seifert-FPM, Germany) under the following conditions: CoKα radiation/Fe filter, 2Θ/Θ goniometer geometry, step mode with 0.05^{o} 2Θ steps, 3s measurement time per step and with the digital processing of resultant data. For measurement and semi-qualitative evaluation, the software packages RayfleX and RayfleX Autoquan (GE Sensing & Inspection Technologies, USA) were used.

IR MEASUREMENTS

Two techniques of IR spectroscopy were used in this study; bromide pellets and diffuse reflectance IR Fourier Transform (DRIFT) spectroscopy.

Approximately 5-10 mg of sample was ground with approximately 400 mg of dried KBr. This mixture was used to collect IR spectra by the DRIFT technique. The IR spectra were collected using the FTIR spectrometer Nexus 470 (ThermoScientific, USA) with a deuterated TriGlycine sulfate (DTGS) detector. The measurement parameters were the following: spectral region 4000-400 cm^{-1}; spectral resolution 8 cm^{-1}; 128 scans; and Happ-Genzel apodization. Freshly dried KBr was used for the background measurement. Every sample was prepared and measured 3-5 times. The mean IR spectrum of every sample was calculated. The mean IR spectra were subsequently used for the creation of chemometric models.

Exactly 0.5 mg of sample was ground with 200 mg of dried KBr. This mixture was used to prepare the bromide pellet. In this study, 13 mm diameter pellets were used. The pellets were pressed by 10 tons for 30 seconds under vacuum. The IR spectra were collected using the FTIR spectrometer Avatar 320 (ThermoScientific, USA) with DTGS detector. The measurement parameters were the following: spectral region 4000-400 cm^{-1}; spectral resolution 8 cm^{-1}; 64 scans; and Happ-Genzel apodization. An empty sample compartment was used for background measurement. Every sample was prepared and measured only once. The IR spectra were subsequently used for the creation of chemometric models.

CHEMOMETRIC ANALYSIS

The chemometric analysis was performed using The Unscrambler 9.7 software package (CAMO Software AS, Norway). PCA and PLS regressions were used as representative of chemometric methods. PCA was used for preliminary data analysis: detecting outlier spectra and the specification of important spectral regions. The PLS1 technique was employed to create chemometric models for the determination of minerals (chlorite, muscovite, albite and quartz) in claystones and clay shales. Multiplicative scatter correction (full MSC) was performed for transformation of DRIFT spectra. The number of optimal PLS parameters were determined by statistical comparison of PRESS values as a function of numbers of factor. The model validation was performed by the cross-validation (CV). The segmented cross-validation was performed as the validation method for calibration models. The size of the cross-validation segment was two samples.

The data matrices of spectral and concentration information of all samples in calibration set were prepared. One data matrix was prepared from IR spectra of bromide pellets; the next data matrix was prepared from DRIFT spectra. Principal component analysis (PCA) was performed for detecting the outlier spectra in the calibration set and for selection of the important spectral regions. The following plots were prepared: the score plot of the first two principal components, the influence plot of the first three

Table 1 List of calibration set of minerals.

Sample	Chlorite (w/w %)	Muscovite (w/w %)	Albite (w/w %)	Quartz (w/w %)
B1	8.2	40.3	3.0	44.9
B2	8.2	42.1	3.3	39.5
B3	8.9	39.0	2.6	46.2
B4	3.5	38.2	3.1	39.1
B5	5.4	8.0	3.5	56.4
B6	9.7	23.5	2.5	64.3
B7	10.5	35.1	1.8	52.6
B8	7.9	48.8	9.2	34.1
B9	6.2	65.5	8.4	20.0
B10	7.8	53.3	4.7	29.4
B11	8.8	56.0	5.6	36.1
B12	5.1	52.1	4.0	33.2
B13	< 1.0	2.2	< 1.0	48.7
B14	3.8	9.7	< 1.0	57.0
B15	< 1.0	16.5	< 1.0	30.6
B16	3.4	17.7	1.5	19.1
B17	< 1.0	1.6	< 1.0	31.4
B18	5.0	12.4	< 1.0	15.7
B19	2.8	17.5	< 1.0	37.2
B20	3.8	10.5	< 1.0	13.5
B21	5.5	14.4	< 1.0	27.3
B22	5.8	14.9	2.3	41.6
B23	< 1.0	10.9	< 1.0	23.1
B24	< 1.0	9.1	< 1.0	23.5
B25	6.0	10.1	1.8	22.1
B26	< 1.0	19.0	2.6	26.6
B27	< 1.0	10.7	1.1	25.2
B28	< 1.0	28.4	< 1.0	28.4
B29	< 1.0	13.4	< 1.0	34.5
B30	1.6	22.6	< 1.0	61.4
B31	< 1.0	9.3	1.7	34.0
B32	5.4	15.4	< 1.0	23.9
B33	22.2	39.5	18.9	19.5
B34	17.6	32.4	19.8	30.1
B35	15.6	27.8	16.8	39.8
B36	9.1	21.6	15.6	53.8
B37	< 1.0	< 1.0	1.8	69.6
B38	< 1.0	34.1	< 1.0	29.2
B39	< 1.0	24.1	< 1.0	6.5
B40	10.4	23.2	18.0	48.4
B41	28.1	41.6	9.1	18.9
B42	17.2	31.1	17.9	26.3
B43	17.2	53.8	3.7	25.2
B44	22.7	25.1	21.4	30.3
B45	17.6	37.2	8.9	22.0
B46	< 1.0	71.9	2.9	5.0
B47	17.0	29.2	14.6	25.8
B48	11.6	47.6	3.8	36.9
B49	29.6	43.3	7.0	20.2
B50	19.2	33.7	19.0	28.1
B51	54.6	12.5	2.5	27.4
B52	3.3	38.1	11.2	43.0
B53	9.6	18.6	41.5	30.4
B54	12.3	25.0	2.3	49.5
B55	20.2	33.3	13.4	32.5
B56	18.0	25.4	14.7	36.9
B57	22.8	43.9	10.6	21.1
B58	25.5	42.5	2.7	27.5
B59	19.8	50.4	1.4	29.4
B60	19.6	30.4	13.2	33.1
B61	20.9	31.1	12.9	35.2
B62	23.4	36.5	12.3	27.8
B63	22.8	34.0	10.9	32.4
B64	13.9	33.8	16.4	35.9
B65	10.8	9.9	27.0	52.3
B66	9.9	42.0	13.5	34.3
B67	17.3	27.3	17.3	32.7
B68	17.6	34.7	13.2	34.5
B69	21.0	44.9	3.0	27.7
B70	22.5	45.9	3.0	25.9
B71	18.9	32.5	15.3	33.4
B72	16.9	23.6	13.0	40.6
B73	21.6	31.9	12.6	34.0
B74	19.5	35.0	16.2	25.0
B75	14.7	31.8	18.3	31.0
B76	16.6	31.3	16.7	33.7
B77	20.5	30.2	17.0	31.3
B78	19.0	29.7	15.8	35.5

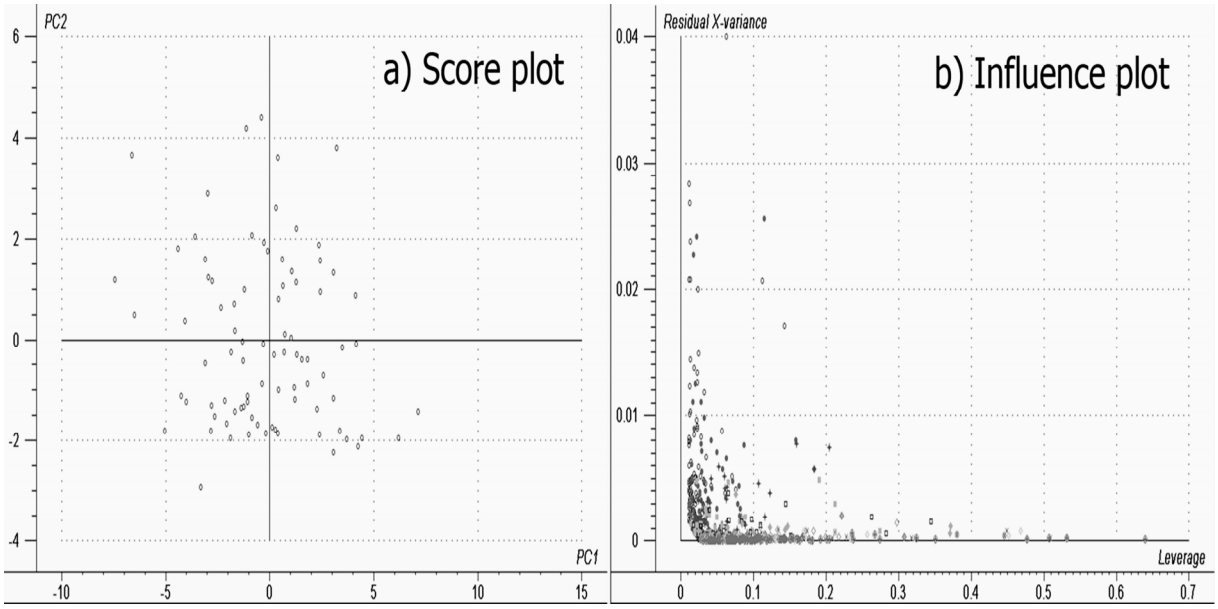

Fig. 2 Score plot (a) and influence plot (b) of DRIFT spectra.

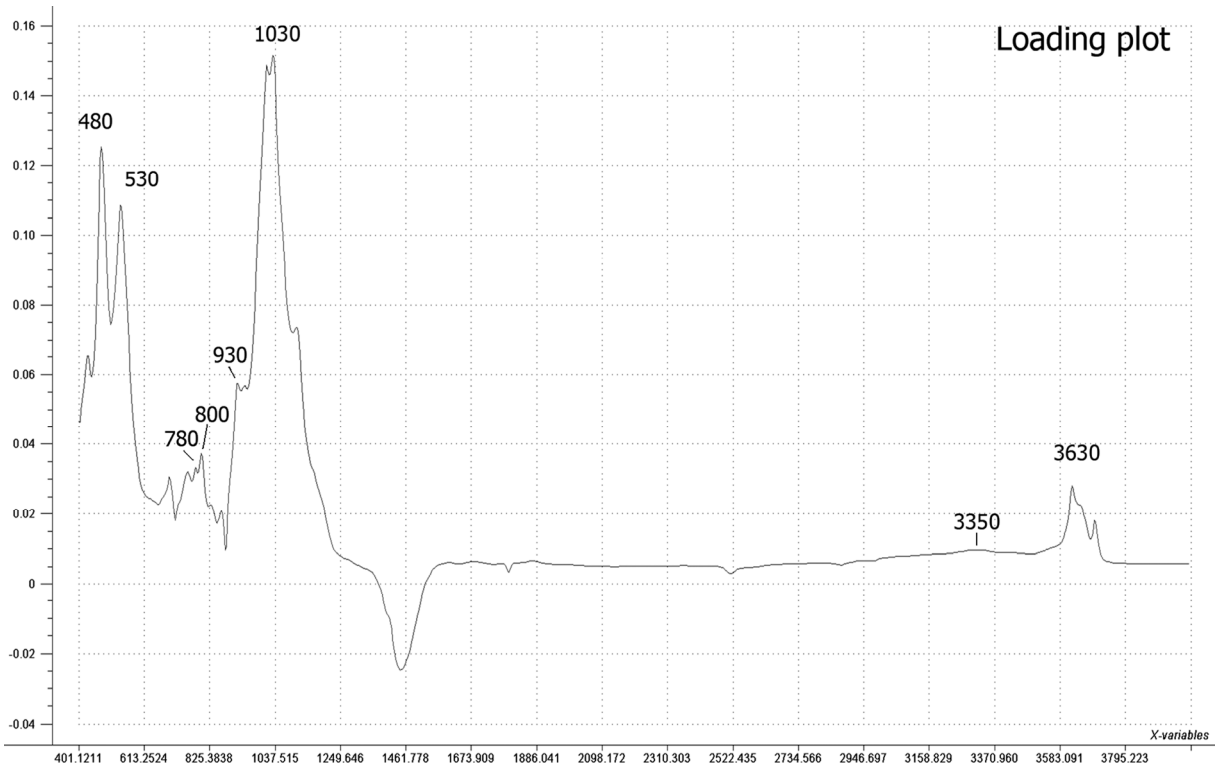

Fig. 3 Line loading plot of bromide pellets spectra.

principal components and the loading plot for the first principal component. The score plots and the influence plots were used to detect the outlier spectra. In the score plot, the outlier spectra are located outside of main cluster. In the influence plot, the outlier spectra do not show decreasing tendency. No outlier spectra were found in either of the data matrices. Examples of both plots for spectra measured by DRIFT technique are shown in Figure 2. The line loading plots were used for the selection of the important spectral regions. This type of loading plot looks like spectrum. Thus the important spectral regions have character of spectral bands. The important spectral regions determined by loading plots of both data matrices were 4000-3000 cm^{-1} and 1300-400 cm^{-1}. The spectral bands present at the spectral region 4000-3000 cm^{-1} belonged to the stretching vibration of structural hydroxyl groups (3630 cm^{-1})

Table 2 Parameters of PLS models.

Mineral	Method	RMSEC (% w/w)	RMSECV (% w/w)	No. of factors	Explained variance (%)
Chlorite	DRIFT	2.67	3.34	8	98.3
	KBr pellet	3.71	4.51	9	96.3
Muscovite	DRIFT	5.58	6.85	8	97.8
	KBr pellet	11.39	11.74	9	95.9
Albite	DRIFT	2.38	2.71	8	98.5
	KBr pellet	2.22	3.44	9	96.5
Quartz	DRIFT	4.91	6.50	11	90.1
	KBr pellet	4.79	6.64	11	87.7

and to the stretching vibration of water (3350 cm^{-1}). The most significant spectral bands in the region 1300-400 cm^{-1} could be assigned to following vibrations: Si-O stretching vibration (1030 cm^{-1}), the deformation vibration of Al-Al-OH (930 cm^{-1}), the Si-O stretching vibrations of quartz (800 cm^{-1} and 780 cm^{-1}) and the deformation vibrations of Al-O-Si and Si-O-Si (530 cm^{-1} and 480 cm^{-1}, respectively). The assignment of spectral bands was performed according to (Ritz et al., 2010). Example of line loading plot for spectra of bromide pellets (including significant wavenumbers) are shown in Figure 3. The "negative" feature in Figure 3 belonged to the vibration of carbonates; this band was not used for creation of calibration models.

The data matrices mentioned previously were used for the creation of the PLS models. The PLS1 technique was used: a separate calibration model was created for determination of each mineral. The important parameters of the PLS models that were created are shown in Table 2. The example of regression between predicted and measured values for PLS model of muscovite (DRIFT spectra) is shown in Figure 4.

The following parameters are shown in Table 2: RMSEC (root mean squared error of calibration), RMSECV (root mean squared error of cross-validation), number of PLS factors and percentage of explained variance. The calibration error of the PLS model was expressed by RMSEC:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(c_{i,cal,pred} - c_{i,refer})^2}{n}} \quad (1)$$

where $c_{i,cal,pred}$ is the value of the mineral content of the i^{th} calibration sample predicted from the PLS model, and $c_{i,refer}$ is the value of the mineral content of the i^{th} calibration sample obtained by the reference method (XRD analysis) and n is the number of samples in the calibration set. The validation of PLS models was performed by the segmented cross-validation. The validation error of the model was expressed by RMSECV, analogous to RMSEC:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(c_{i,val,pred} - c_{i,refer})^2}{n}} \quad (2)$$

where $c_{i,val,pred}$ is the value of the mineral content of the i^{th} validation sample predicted by the PLS model, and $c_{i,refer}$ is the value of the mineral content of the i^{th} validation sample obtained by the reference method and n is the number of samples in the calibration set. The number of factors is the optimal number of "latent variables" necessary to effectively describe the PLS model. The number of factors was obtained by so called PRESS plot (i.e., the plot of PRESS vs. the number of factors). PRESS mean predicted residual error sum of squares and this parameter was calculated as:

$$PRESS = \sum_{i=1}^{n}(c_{i,pred} - c_{i,refer})^2 \quad (3)$$

where $c_{i,pred}$ is the value of the mineral content of the i^{th} sample predicted from the PLS model, and $c_{i,refer}$ is the value of the mineral content of the i^{th} sample obtained by the reference method and n is the number of samples in the calibration set. The explained variance is the percentage of the variance of the system described by the expressed number of factors.

RESULTS AND DISCUSSION

ANALYSIS OF CONTROL SAMPLES (ACCURACY AND PRECISION)

The predictive ability of the PLS models was tested by analysis of the eight control samples (CS1-CS8). The KBr pellets were prepared for each control sample and the IR spectra of these control samples were subsequently measured. The IR spectra of the control samples were also obtained by the DRIFT technique (each spectrum was prepared like the mean spectrum from three independent DRIFT measure-

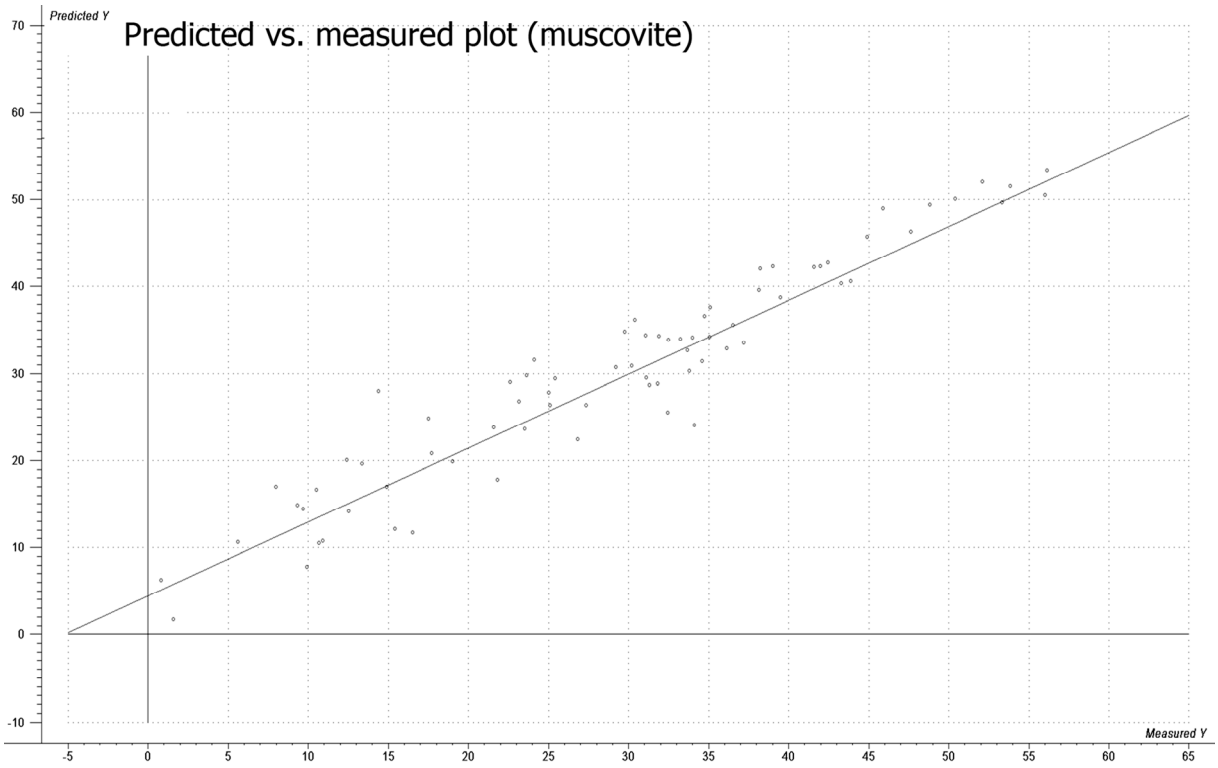

Fig. 4 Predicted vs. measured plot (Muscovite PLS model; DRIFT spectra).

Table 3 List of control samples and results of their analysis.

	Chlorite (% w/w)			Muscovite (% w/w)			Albite (% w/w)			Quartz (% w/w)		
	XRD	PLS		XRD	PLS		XRD	PLS		XRD	PLS	
Sample		DRIFT	Pellets		DRIFT	Pellets		DRIFT	Pellets		DRIFT	Pellets
CS-1	10.7	8.1	7.5	12.1	16.2	10.9	4.5	5.6	3.9	31.6	23.9	26.2
CS-2	3.0	5.0	2.6	37.4	39.8	42.3	6.7	4.8	6.2	52.9	47.2	47.9
CS-3	5.0	5.2	5.0	40.8	43.2	37.7	8.2	4.6	7.0	48.9	46.1	44.9
CS-4	14.8	12.4	13.0	87.5	59.2	33.8	5.9	7.3	5.8	14.9	15.8	19.5
CS-5	19.1	19.9	18.7	36.5	37.0	36.3	14.3	13.6	15.1	30.0	30.4	32.9
CS-6	24.3	20.8	23.8	30.9	29.8	33.5	14.5	15.2	14.2	29.4	28.3	29.4
CS-7	11.7	15.5	10.2	30.0	27.1	29.0	41.5	31.8	47.0	30.4	32.5	26.8
CS-8	21.0	20.3	16.5	27.2	30.2	25.1	12.0	12.3	10.4	39.8	33.3	34.4

ments). All of these spectra were used for prediction of the content of minerals by the PLS models that were created. The results of the analysis of control samples from PLS models are shown in Table 3 together with the content of minerals in the control samples obtained from the reference method (XRD analysis).

First, the results of prediction of the content of minerals in control samples were tested for statistical compliance with the reference values (results of XRD analysis) of the control samples. The testing techniques were the following: F-test, t-test (Student's test) and paired comparison (Meloun and Militký, 2004). All of these techniques showed statistical compliance between the predicted and reference values for the control samples.

The predictive ability of chemometric models can be described using several validation diagnostics. The following parameters were used in this study: bias, standard error of prediction (SEP) and mean relative error (RE). Bias and SEP parameters were used according (Esbensen, 2006); RE parameters was created for the purpose of this study:

$$bias = \frac{\sum_{i=1}^{n}\left(c_{i,pred} - c_{i,refer}\right)}{n} \tag{4}$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}\left(c_{i,pred} - c_{i,refer}\right)^2}{n}} \tag{5}$$

Table 4 Parameters of predicted ability of PLS models.

	Chlorite		Muscovite		Albite		Quartz	
Parameter	DRIFT	Pelets	DRIFT	Pelets	DRIFT	Pelets	DRIFT	Pelets
bias (w/w %)	-0.3	-1.5	-2.5	-6.7	-0.1	0.1	-1.9	-1.2
SEP (w/w %)	2.4	2.1	10.3	19.1	1.7	1.6	4.9	4.2
RE (%)	20.7	11.5	12.9	13.9	19.4	9.1	10.5	12.6

Table 5 Reproducibility - list of results.

	Chlorite (w/w %)				Muscovite (w/w %)				Albite (w/w %)				Quartz (w/w %)			
	CS5		B25		CS5		B25		CS5		B25		CS5		B25	
	DRIFT	Pellets	DRIFT	Pellets	DRIFT	Pellets	DRIFT	Pellets	DRIFT	Pellets	DRIFT	Pellets	DRIFT	Pellets	DRIFT	Pellets
	18.8	17.3	2.5	6.2	38.0	28.4	12.7	11.5	11.9	11.4	1.4	2.6	32.7	32.2	27.6	23.9
	19.0	21.5	2.2	6.1	37.6	30.9	12.0	12.8	12.3	14.2	1.7	1.0	31.9	30.6	27.6	23.1
	19.3	17.1	2.3	5.2	38.1	27.1	11.7	10.8	12.1	14.3	1.9	1.5	31.6	33.5	27.3	25.3
	18.5	19.4	2.1	6.3	36.2	32.3	11.9	11.8	11.8	12.4	1.5	1.3	31.7	33.3	27.7	25.7
	18.8	18.7	2.2	4.5	37.3	27.4	11.8	11.0	11.9	14.0	1.6	1.6	31.9	33.2	27.2	22.1
	18.8	20.7	2.3	6.9	37.0	28.6	12.0	11.5	11.8	12.7	1.3	3.2	31.2	34.0	27.6	24.5
	18.5	18.7	2.1	6.7	37.1	25.3	10.9	10.1	11.4	11.7	1.6	2.5	32.4	30.8	28.2	25.7
	18.7	17.2	2.7	5.6	38.0	27.9	9.7	9.4	11.2	13.6	1.7	3.2	32.7	34.4	28.1	25.5
	18.5	18.0	3.0	5.5	37.2	26.8	10.3	12.2	11.4	14.1	1.4	1.1	32.7	32.0	27.6	22.3
	18.9	17.5	2.1	5.6	37.9	28.5	11.9	10.7	11.6	13.2	1.5	2.2	32.8	32.6	27.8	23.3
RSD (%)	1.4	8.3	12.8	12.1	1.6	7.1	7.9	8.8	3.0	9.2	10.9	40.5	1.7	4.0	1.1	5.8

$$RE = \frac{\sum_{i=1}^{n}\left(\frac{\left|c_{i,pred} - c_{i,refer}\right|}{c_{i,refer}}\right)}{n} \cdot 100 \qquad (6)$$

where $c_{i,pred}$ is the value of the mineral content of the i[th] control sample predicted from the PLS model, $c_{i,refer}$ is the value of the mineral content of the i[th] control sample obtained by the reference method (XRD) and n is number of control samples.

Bias represents the average difference between the predicted values and the reference values for control samples and is a commonly-used measure of the accuracy of a chemometric model. Bias is also used to check any systematic differences observed between the average values of the control samples and the validation samples (Esbensen, 2006). Another way to express the accuracy of a chemometric model is the mean relative error. The standard error of prediction (SEP) expresses the precision of the predicted results. The values of the parameters of the validation diagnostics are shown in Table 4.

The bias of almost all of the PLS models was negative; only the bias of the PLS model of albite for KBr pellets had a positive value. The absolute values for bias were usually very low, ranging from 0.1 to 1.9. Only the bias of the models of muscovite showed higher values; the model for DRIFT had a bias of -2.5 and the model for the KBr pellets had a bias value of -6.7. These high values of bias were very probably caused by the worse values of the parameters of the muscovite PLS models (see Table 2).

The accuracy as expressed by the bias showed worse values for the muscovite PLS models, whereas the accuracy of the muscovite PLS models as expressed by the RE showed values similar to the RE of the rest of models. The reason is obvious from the different mathematical formulas for bias and RE (see equations 4 and 6). The bias equation (4) does not operate with the absolute values of the difference predicted and the reference content of the mineral. Negative and positive increments could therefore cancel each other. The RE equation (6) used absolute values of the difference. RE is therefore the more robust parameter for accuracy, whereas bias can describe the systematic error of the PLS model. The accuracy parameter (RE) showed similar values in most of the models (approximately 10 %). Only DRIFT models for the prediction of chlorite and albite had values of RE of approximately 20 %. The values of the accuracy for the PLS models that were created are very similar to the values of accuracy for the results of the XRD analysis cited in the literature (e.g. Moore and Reynolds, 1997). Requirement of practice for RE parameter of quantitative phase analysis results of rock are about 20-25 %. The values of RE parameters of control samples achieved by reference quantitative method (XRD analysis) in this study were between 10 % and 20 %. Thus both methods (XRD analysis and chemometric analysis of IR spectra) provided results acceptable by potential customers.

The best values for the parameter of precision (SEP) had PLS models for the prediction of albite and chlorite (values of SEP about 2 w/w %). The models for prediction of quartz showed slightly worse precision. The model for the prediction of muscovite had dramatically worse precision than other PLS models, probably caused by the worse parameters of the muscovite PLS models (see Table 2). With the

exception of the models for the prediction of muscovite, the other models had very similar values of SEP for DRIFT and KBr pellet spectra. The precision of analysis of control samples by reference quantitative method (XRD analysis) in this study were between 10 % and 20 %.

REPRODUCIBILITY

One sample from the calibration set (B25) and one control sample (CS5) were used to estimate the reproducibility of the PLS models that were created. For three weeks, ten KBr pellets were prepared from each sample, and their IR spectra were measured. In the same period, ten IR spectra were also obtained by the DRIFT technique; each spectrum was prepared like the mean spectrum from three independent DRIFT measurements (including homogenization and grinding with KBr). All of these spectra were used for the prediction of the content of minerals by the PLS models that were created. The reproducibility was expressed by the relative standard deviation (RSD). RSD was calculated from the results that were obtained:

$$RSD = \frac{\sqrt{\sum_{i=1}^{n}(x_P - x_i)^2}}{\frac{n-1}{x_P}} \cdot 100 \qquad (11)$$

where x_i is the predicted value of the mineral content of the ith analysis of reproducibility, x_P is the mean value of the predicted mineral content and n is number of analyses. The results and the calculated RSD are shown in Table 5.

Worse values of reproducibility were obtained from the analysis of the IR spectra using KBr pellets. Values of the RSD in the analyses of chlorite, muscovite and albite were approximately 10 %; the values of the RSD in the analysis of quartz were approximately 5 %. The extremely high value of the RSD of sample BP5 (in the analysis of albite) was caused by the very low content of albite in this sample.

Significantly better values of the RSD were obtained from the analysis of the IR spectra measured by the DRIFT technique. Most of values of the RSD were within the range 1-3 %. The higher values of the RSD (about 10 %) were obtained from the analysis of samples with a low content of minerals (chlorite, muscovite and albite in sample BP5).

Different values of the reproducibility from the analysis of DRIFT spectra and the spectra from KBr pellets were very probably caused by the different weights of samples in preparing for IR measurement. For preparation of DRIFT measurements, 5-10 mg of the samples was used, whereas for preparing KBr pellets, only 0.5 mg of sample was used. This relatively low weight is at the limits of accurate measurement for common analytical balances and probably caused an elevated error of weighing. However, the use of a heavier sample could induce total absorbance of some spectral bands and cause poorer utilization of spectral information.

CONCLUSIONS

Principal component analysis (PCA) and especially PLS regression were used as chemometric methods in this paper. PCAs were used for detecting outliers and for selection of important spectral regions. No outliers were detected. Regions 4000-3000 cm^{-1} and 1300-400 cm^{-1} were selected for the creation of calibration models by the PLS regression technique. The series of PLS models proposed in this work showed RMSEC and RMSECV values up to 5-6 w/w % (with the exception of the PLS model for prediction of muscovite based on the KBr pellet technique). For a set of control samples, values of mean relative error (RE) of about 10 % were achieved in most of the PLS models that were created. The slightly better values of RE were achieved for PLS models based on the KBr pellet technique. The reproducibility of the PLS models was evaluated for two samples and the reproducibility was expressed by the relative standard deviation (RSD); the values of the RSD ranged from 1.1 to 12.8 % (with the exception of the PLS model for prediction of albite based on the DRIFT technique). The relatively better values of the RSD were achieved for the PLS models based on the DRIFT technique.

There are some important advantages of chemometric analysis of IR spectra over most used quantitative phase analysis method (Rietveld technique of XRD analysis): analysis time, accessibility and simplicity. First, the analysis time of chemometric analysis of IR spectroscopy is much shorter than analysis time of XRD analysis. Of course, creation of calibration models is rather time-consuming. But subsequent analysis is very fast. Second, IR spectrometers are present in significantly larger numbers of laboratories than X-ray diffractometer. The main reasons are the purchase price of IR spectrometers and operating costs, which helped to spread the use of infrared spectroscopy as a common analytical technique. And finally, chemometric result processing of IR spectra is considerably simpler than Rietveld treatment of diffractogram data.

The use of chemometric data treatment for the IR spectra is simple and reliable and can be used to determine minerals in rocks. This study showed that IR spectroscopy in conjunction with the PLS regression method provided an acceptable alternative to the most commonly used methods for quantitative phase analysis – the Rietveld technique of XRD analysis.

ACKNOWLEDGMENTS

This paper was created by the project No. CZ.1.05/2.1.00/01.0040 "Regional Materials Science

and Technology Centre" within the framework of the operation program "Research and Development for Innovations" financed by the Structural Funds and from the state budget of the Czech Republic.

This paper also has been done in connection with project Institute of clean technologies for mining and utilization of raw materials for energy use, reg. no. CZ.1.05/2.1.00/03.0082 supported by Research and Development for Innovations Operational Programme financed by Structural Founds of Europe Union and from the means of state budget of the Czech Republic.

REFERENCES

Al-Degs, Y.S, El-Sheikh, A.H., Al-Ghouti, M.A., Hemmateenajed, B. and Walker, G.M.: 2008, Solid-phase extraction and simultaneous determination of trace amounts of sulphonated and azo sulphonates dyes using microemulsion-modified-zeolite and multivariate calibration. Talanta 75, 904–915.

Armenta, S., Garrigues, S. and de la Guardia, M.: 2007, Determination of edible oil parameters by near infrared spectroscopy. Anal. Chim. Acta, 596, 330–337.

Bish, D.L. and Howard, S.A.: 1988, Quantitative phase analysis using the Rietveld method. J. Appl. Crystallography, 21, 86–91.

Breen, C., Cleeg, F., Herron, M.M., Hild, G.P., Hillier, S., Hughes, T.L., Jones, T.G.J., Matteson, A. and Yarwood, J.: 2008, Bulk mineralogical characterization of oilfield reservoir rocks and sandstones using Diffuse Reflectance Infrared Fourier Transform Spectroscopy and Partial Least Square analysis. J. Petrol. Sci. Eng., 60, 1–17.

Brindley, G.W.: 1980, Crystal structures of clay minerals and their X-ray Identification. Mineralogical Society, London.

Chipera, S.J. and Bish, D.L.: 2001, Baseline studies of the clay minerals society source clays: Powder X-ray Diffraction Analyses. Clays and Clay Minerals, 49 (5), 398–409.

Esbensen, K.H.: 2006, Multivariate data analysis in practice. Fifth ed., Camo, Oslo.

Fredericks, P.M., Lee, J.B., Osborn, P.R. and Swinkels, D.A.J.: 1985, Material characterization using factor analysis of FT-IR Spectra. Part 1: Results. Appl. Spectrosc. , 39 (2), 303–316.

Fuller, M.P., Ritter, G.L., Draper, C.S.: 1988, Partial Least-Squares quantitative analysis of infrared spectroscopic data. Part I: Algorithm Implementation. Appl. Spectrosc. , 42 (2), 217–236.

Geladi, P. and Kowalski, B.R.: Partial Least-Squares regression: A tutorial. Anal. Chim. Acta, 185 (1986), 1–17.

Haaland, D.M. and Thomas, E.V.: 1988, Partial Least-Squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of Qualitative Information. Anal. Chem., 60, 1202–1208.

Hasegawa, T.: 2002, Principal component regression and partial least quares modelig. In Chalmers, J. M. and Griffits, P.R. (eds.), Handbook of Vibrational Spectroscopy, 3, John Wiley and Sons, Chichester.

Hillier, S.: 2000, Accurate quantitative analysis of clay and other minerals in sandstones by XRD: comparison of a Rietveld and a reference intensity ratio (RIR) method and the importance of sample preparation. Clay Minerals, 35 (1), 291–302.

Iñón, F.A., Garrigues, J.M., Garrigues, S., Molina, A. and de la Guardia, M.: 2003, Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance-Fourier transform infrared spectroscopy. Anal. Chim. Acta, 489, 59–75.

Kodama, H., Kotlzar, L.S. and Ripmeester, J.A.: 1989, Quantification of crystalline and noncrystalline material in ground kaolinite by X-ray powder diffraction, infrared, solid-state nuclear magnetic resonance, and chemical-dissolution analyses. Clays and Clay Minerals, 37 (4), 364–370.

Lorber, A. and Kowalski, B.R.: 1988, A note on the use of the partial Least-Squares method for multivariate calibration. Appl. Spectrosc., 42 (8), 1572–1574.

Luis, M.L., Fraga, J.M.G., Jimenéz, A.I., Jimenéz, F., Hernandéz, O. and Arias, J.J.: 2004, Application of PLS regression to fluorimetric data for the determination of furosemide and triamterene in pharmaceutical preparations and triamterene in urine. Talanta 62, 307–316.

Martens, H. and Naes T.: 1989, Multivariate Calibration. John Willey and Sons, New York.

Meloun M. and Militký, J.: 2004, Statistical analysis of experimental data. Second ed., Academia, Praha, (in Czech).

Moneeb, M.S.: 2006, Polarographic chemometric determination of zinc and nickel in aqueous samples. Talanta, 70, 1035–1043.

Moore, D. M. and Reynolds, R. C.:1997, X-Ray diffraction and the identification and analysis of clay minerals. Oxford University Press, New York.

Rietveld, H.M.: 1969, A profile refinement method for nuclear and magnetic structures. J. Appl. Cryst., 2, 65–71.

Ritz, M. Vaculíková, L. and Plevová, E.: 2010, Identification of clay minerals by infrared spectroscopy and discriminant analysis. Appl. Spectrosc., 64 (12). 1379–1387.

Srodon J.: 2002, Quantitative mineralogy of sedimentary rocks with emphasis on clays and with applications to K-Ar dating. Mineralogical Magazine, 66 (5), 677-687

Vogt, Ch., Lauterjung, J. and Fischer, R.X.: 2002, Investigation of the Clay Fraction (<2 μm) of the Clay Minerals Society Reference Clays. Clays and Clay Minerals, 50 (3), 388–400.

Wagieh, N.E., Hegazy, M.A., Abdelkawy, M., Abdelaleem, E.A.: 2010, Quantitative determination of oxybutynin hydrochloride by spectrophotometry, chemometry and HPLC in presence of its degradation and additives in defferent pharmaceutical dosage forms. Talanta, 80, 2007–2015.

Wold, S., Sjöström, M. and Eriksson, L.: 2001, PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst., 58, 109–130.

Zapata-Urzúa, C., Pérez-Ortiz, M., Bravo, M., Olivieri, A.C. and Álvarez-Lueje, A.: 2010, Simultaneous voltametric determination of levodopa, carbidopa and benserazide in pharmaceuticals using multivariate calibration. Talanta, 82, 962–968.